Localization of the Human Stat6 Gene to Chromosome 12q13.3–q14.1, a Region Implicated in Multiple Solid Tumors

Bharvin K. R. Patel, Catherine L. Keck,* Ryan S. O'Leary, Nicholas C. Popescu,* and William J. LaRochelle¹

Laboratory of Cellular and Molecular Biology and *Laboratory of Experimental Carcinogenesis, National Cancer Institute, Building 37 Room 1E24, Bethesda, Maryland 20892

Received March 9, 1998; accepted June 3, 1998

Stat6 signaling pathways have been correlated with functional responses induced by IL-4 and PDGF that may play a role in human malignancy. Utilizing fluorescence *in situ* hybridization, we mapped the human Stat6 gene to chromosome 12q bands 13.3-14.1, a breakpoint region implicated in a wide variety of solid tumors. To understand the genesis of three human Stat6 variant cDNAs, including a naturally occurring dominant negative species, we further characterized the genomic structure and flanking regions of the human Stat6 gene. The human Stat6 gene encompassed over 19 kb and contained 23 exons. For promoter studies, we introduced flanking sequence 5' of Stat6 exon 1 into a promoterless luciferase reporter vector and characterized basal promoter activity by deletion analysis. DNA sequence analysis revealed potential transcriptional regulation of the putative promoter through numerous consensus binding elements. Finally, we conclude that selective exon deletion and utilization of alternative donor/ acceptor sites appear to explain best human Stat6 variant mRNAs. © 1998 Academic Press

INTRODUCTION

The Stats (signal transducers and activators of transcription) are a family of transcription factors (Darnell *et al.*, 1994; Ihle and Kerr, 1995; Schindler and Darnell, 1995) that have been evolutionarily conserved from *Drosophila* to humans. Six murine Stat loci have been mapped to three clusters on different autosomes, suggesting that in mouse the Stats have arisen by tandem duplication of the ancestral locus followed by dispersion of the linked loci to different chromosomes (Copeland *et al.*, 1995). Physiologically, Stat signaling pathways have been correlated with pleiotropic functional responses induced by a variety of growth factors, cytokines, and in-

¹ To whom correspondence should be addressed. Telephone: (301) 496-9052. Fax: (301) 496-8479. E-mail: billr@helix.nih.gov.

terferons (Darnell *et al.*, 1994; Schindler and Darnell, 1995; Shuai *et al.*, 1994). Selective Stat activation, presumably through tyrosine phosphorylation and subsequent dimerization, results in its translocation to the nucleus (Shuai *et al.*, 1994). Transcriptional activation occurs in part after Stat binding with varying affinity to unique DNA motifs related to interferon γ -activated sites. Therefore, selective differences in DNA site selection and Stat expression appear to determine the transcriptional consequences and subsequent proliferation or differentiation events that occur in a given cell in response to this activation (Schindler *et al.*, 1995).

Stat6 (Hou *et al.*, 1994; Schindler *et al.*, 1994; Quelle *et al.*, 1995) activation correlates with functional responses induced by IL-4 (Pernis *et al.*, 1995), IL-13 (Malabarba *et al.*, 1996), and PDGF (Patel *et al.*, 1996). Phenotypic analysis of Stat6^{-/-} mice have elegantly demonstrated a role for Stat6 in IL-4- and IL-13-induced lymphocyte proliferation, Th2 helper T cell (Seder *et al.*, 1992) differentiation, immunoglobulin class switching, and MHC class II and CD23 cell surface antigen expression (Kaplan *et al.*, 1996; Shimoda *et al.*, 1996; Takeda *et al.*, 1996). However, the mechanism(s) by which Stat6 induces these effects as well as the role of Stat6 in PDGF-induced proliferation and transcriptional activation remains incompletely understood.

PDGF is a proliferative and chemotactic factor for mesenchymal cells (Aaronson, 1991; Heldin and Westermark, 1991). Although PDGF plays an important role in normal development and wound repair, accumulating evidence suggests that its abnormal expression contributes to a variety of fibrotic diseases (Smits et al., 1992) and mesenchymal tumors (Fleming et al., 1992; Nister et al., 1991). PDGF-induced proliferation as well as Stat6 and Jak1 activation in NIH 3T3 fibroblasts is enhanced by IL-4, strongly suggesting that IL-4 might affect some PDGF-mediated biological responses in vivo (Patel et al., 1996). Recently, we reported that human tissues contain three alternatively spliced variants of Stat6, including attenuated and dominant negative species encoded by unique cDNAs (Patel et al., 1998). Since these variants represent alternative transcriptional processes by which

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. AF067572, AF067573, AF067574, and AF067575.

HUMAN Stat6 GENE MAPPING AND GENOMIC STRUCTURE



FIG. 1. Exon–intron organization of the human Stat6 gene. Solid black boxes represent exons and horizontal lines indicate introns. Restriction endonuclease cleavage sites determined from individual λ clones are depicted.

Stat6 activity could be regulated, we here isolated and determined the genomic structure of the human Stat6 gene. To understand the differential regulation and splicing of these variants, we characterized the human Stat6 basal promoter. We further evaluated the potential role of Stat6 in human disease after chromosomal mapping.

MATERIALS AND METHODS

Isolation, characterization, and DNA sequence analysis of the human Stat6 gene. A human placental genomic library packaged in the Lambda FIX II vector (Stratagene, La Jolla, CA) was screened with a ³²P-labeled Stat6 cDNA probe using standard recombinant DNA protocols (Beeler *et al.*, 1997). Positive plaques were isolated and purified by secondary screening. The human Stat6 genomic clone (pBP95) was also isolated from a human P1 genomic library (Genome Systems, Inc., St. Louis, MO). Regions of the P1 clone were used to confirm and span gaps among our Lambda FIX II genomic clones. DNA was purified from positive plaques and mapped by restriction enzyme digestion combined with the Southern blot analysis. Fragments that hybridized to the Stat6 cDNA probe were isolated from our $\lambda 3/\lambda 4$ clones and subcloned into pBluescript II KS(⁺) (Stratagene, Palo Alto, CA) to yield the plasmids pBP2, pBP3, pBP68, and pBP74. Both cDNA strands were completely sequenced using an ABI PRISM dye terminator cycle sequencing kit and automated ABI PRISM 310 genetic analyzer or manually by Sequenase kit (United States Biochemical, Cleveland, OH) using human Stat6 cDNA sequence-derived primers. GenBank accession numbers for the human Stat6 genomic clone are AF067572, AF067573, AF067574, and AF067575.

193

RNase protection assay. The 678-bp *Eco*RI–*Nhe*I genomic fragment (Fig. 3) was cloned into the *Eco*RI–*Xba*I site of pGEM-3Zf(–). The construct was linearized with *Bsu*36I and used to synthesize the radiolabeled antisense RNA probe by SP6 RNA polymerase in the presence of [³²P]UTP. RNase protection assay was performed using the Ambion RPA II kit (Austin, TX). Labeled RNA probe (5 × 10⁵ cpm) was hybridized to 10 μ g of Ramos poly(A)⁺ RNA or 20 μ g of control yeast RNA. The hybridized products were digested at 37°C for 30 min with a mixture of RNase A (20 units) and RNase T1 (1 μ g) according to the manufacturer's specifications. Protected fragments were resolved on a 6% polyacrylamide/urea gel and visualized by autoradiography.

| Exon Number | Exon | Number of amino acids | Intron and its size (bases) |
|----------------|-------------------------|-----------------------|-----------------------------|
| 1 | 1-AGCCAG GCCGAG-229 | 0 | gtgagg ggctag (1021) |
| 2 | 230-AGAAAG GCTGGG-297 | 0 | gtaagt ctccag (1744) |
| 3 | 298-GCAACC GCCCTG-434 | 38 2/3 | ataagt tacaag (414) |
| 4 | 435-GGAGTT CTTGAG-573 | 46 | gtgggg ctgtag (290) |
| 5 | 574-AGCATA GAACAG-657 | 28 | gtattg gtgtag (399) |
| 6 | 658-TTCCGC GCCAAG-796 | 46 1/3 | gtagag cctcag (112) |
| 7 | 797-TGTCTC AGTGAG-849 | 17 | gtgagt ccatag (187) |
| 8 | 850-GCCCTG GGAGAG-998 | 49 2/3 | gttggg ccccag (592) |
| 9 | 999-GTGTGA CACCAG-1130 | 43 2/3 | gtatte teccag (128) |
| 10 | 1131-TTGCTT TGGAGC-1319 | 62 2/3 | gtaage gggcag (338) |
| 11 | 1320-AGAAAG AACCTG-1407 | 29 | gtgagg ccacag (139) |
| 12 | 1408-CTTCTC CTCCAG-1530 | 41 | gtgaac cttcag (~1450) |
| 13 | 1531-GCCCTG GAGATG-1623 | 31 | qtqaqq ccccaq (334) |
| 14 | 1624-GACCGC AACAAG-1830 | 69 | gttcag ccctag (~4150) |
| 15 | 1831-GAGATC TGACCG-1925 | 31 2/3 | gtgagt ggccag (92) |
| 16 | 1926-GCTGAT AGGATG-2062 | 45 1/3 | gtgagg ccatag (325) |
| 17 | 2063-GCTCTC ACAAGC-2209 | 48 1/3 | gtgage gaacag (227) |
| 18 | 2210-CTGAAC GGAAAG-2273 | 20 2/3 | gtgagt ctccag (112) |
| 19 | 2274-GGACCA TATGGT-2384 | 36 2/3 | gtaagg ccccag (194) |
| 20 | 2385-GCCCCA CCAGGA-2477 | 30 2/3 | gtaagt ttccag (1375) |
| 21 | 2478-GCCTCA CCCCAG-2544 | 22 | gtgaat tggcag (89) |
| 22 | 2545-GGCCTG CACTTG-2672 | 42 2/3 | gtgagt ttgcag (88) |
| 23 | 2673-GATTGG AATCTG-4018 | 62 | aaaagaaaga (>300) |

TABLE 1 Exon-Intron Boundaries of the Human Stat6 Gene

Note. Sizes of exons and introns as well as nucleotide sequences surrounding the exon-intron boundaries are indicated.

Promoter constructs and basal luciferase activity. The KpnI– NheI fragment (~5.5 kb) that contained 134 bp of the first exon and the upstream flanking sequence of Stat6 was isolated from pBP3 and cloned into compatible polylinker sites of the luciferase reporter pGL3-Basic (Promega, Madison, WI), designated pBP78. The pBP78 plasmid construct was double digested with KpnI and BlnI, EcoRI, SpeI, or BssHII, blunted with Klenow polymerase, and self-ligated to engineer pBP82 (-450 to +134), pBP84 (-543 to +134), pBP86 (-947 to +134), and pBP88 (-2454 to +134), respectively.

To study basal promoter activity, murine NIH 3T3 fibroblasts were cultured in six-well plates (3 \times 10⁵ cells per well) for 24 h and then transiently transfected with individual Stat6 promoter-luciferase chimeric constructs using a calcium phosphate precipitation protocol (Patel et al., 1998). Relative luminescent units were normalized after measuring protein concentrations by the method of Bradford. For transfection, 1 μ g of pRL-CMV was used per 3 \times 10⁵ cells. After incubation of the cells with the DNA-calcium phosphate precipitates for 5 h in Dulbecco's modified Eagle's medium (DMEM)/10% calf serum, the cells were washed twice with DMEM and incubated for 72 h in DMEM/10% calf serum before harvesting. For the determination of luciferase activity, the cells were washed and extracted in lysis buffer, and luciferase activity was assayed according to the manufacturer's protocol (Promega). A Lumat-LB luminometer (Berthold) equipped with a dual injector was used to measure luciferase activity.

Fluorescence in situ hybridization. An 8-kb fragment derived from our Stat6-P1 genomic clone was isolated and labeled with biotin–dUTP or digoxigenin-11–dUTP (Random Primed DNA Labeling Kit, Boehringer Mannheim). The labeled probe was used for *in situ* hybridization of human chromosomes derived from methotrexate-synchronized normal peripheral lymphocyte cultures. The conditions of the hybridization, detection of hybridization signals, digitalimage acquisition, processing, and analysis as well as the procedure for direct visualization of fluorescent signals to banded chromosomes were performed as previously described (Popescu *et al.*, 1994; Zimonjic *et al.*, 1995).

RESULTS

Genomic Structure of the Human Stat6 Gene

We have recently cloned and characterized three human Stat6 cDNA species encoding wildtype, attenuated, and dominant negative gene products (Patel et al., 1998). To reveal the processes by which these human Stat6 transcripts could be regulated, we first sought to elucidate the human Stat6 genomic structure. Two nonoverlapping λ DNA clones, designated λ 3 and $\lambda 4$, of the human Stat6 gene were isolated from a Lambda FIX II human genomic DNA library using full-length human Stat6 cDNA as a probe. As shown in Fig. 1, a major portion of the λ 3 clone contained the 5' flanking region of the Stat6 gene, while its downstream region covered exons 1 through 11. The λ 4 clone consisted of exons 15 through 23 and the 3' flanking region downstream of the Stat6 gene. To map exon 12 through 14, an additional genomic clone was isolated by screening a human P1 library with a full-length Stat6 cDNA probe. The P1-Stat6 plasmid clone contained the entire human Stat6 gene, including both the 5' and the 3' flanking regions of the gene as determined by restriction endonuclease and DNA sequence analysis.

Comparison of our genomic clones with the longest Stat6 variant cDNA sequence revealed that the human Stat6 gene was approximately 19 kb in length and



FIG. 2. RNase protection assay of human Stat6 mRNA adjacent to 5' flanking genomic sequence. Human Ramos poly(A)⁺ mRNA was prepared and hybridized with a 497-bp ³²P-labeled *Bsu*36I–*Nhe*I genomic fragment containing the 5' flanking region of the human Stat6 gene adjacent to and including known cDNA encoded sequence. After hybridization, Ramos (lane 1) or yeast (lane 2) transcripts were digested as described under Materials and Methods. Protected bands were visualized after autoradiography. The ³²P-*Bsu*36I–*Nhe*I probe was also subjected to electrophoresis (lane 3). +1 denotes the 134-base protected fragment (+1 also indicated in Fig. 3). Marker positions were determined from the base spacing of a sequenced cDNA resolved in parallel.

consisted of 23 exons interrupted by 22 introns. The exon-intron junction sequences as well as the lengths of the human Stat6 gene's exons and introns are summarized in Table 1. The first 2 exons of the Stat6 gene were noncoding. Most exons were relatively small in size and ranged from 53 (exon 7) to 205 bp (exon 14) with the exception of exon 23. Although exon 23 was the largest in size (1346 bp), it encoded only the carboxy-terminal 62 amino acids of Stat6. Intron sizes varied from 88 (intron 22) to approximately 4150 bp (intron 14). While the sizes of most introns were determined by DNA sequence analysis, the lengths of introns 12 and 14 were determined by PCR amplification using two sets of Stat6-specific primers that flanked these introns. All introns were flanked by the classical GT/AG sequence that conformed well to the consensus for 5' splice donor and 3' splice acceptor sites found in other human genes (Table 1).

Analysis of the distribution of the various motifs encoded by the human Stat6 gene showed that the

HUMAN Stat6 GENE MAPPING AND GENOMIC STRUCTURE

| CTGTTGCCCAGGCTGAAGTGCAAGGGTGTAATCTCAGCTCACTGCAGCCTCTGCCTCCCGGATTCCAGTGATTTTCCTGCTTCAGCTTCCCAGGTAGCTG | -3501 |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|
| GGATTGCAGGCACATGCCACTGTGCCCAGATAATTTTTTGTATTTTTAGTAGAGACGGGGCTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTAACC | -3401 |
| TCAGATGATCCGCACACCTCGGCCTCCCAAAGTGCTAGGATTACAGGCATGAGCCTCTGCGCCTTGGCCTTTACTTCAGTTTCAAAGAGTGAAGGTCAAGA | -3301 |
| CACTGTTTTTTTTTTTGTTTTGTTTTTAATTTTGTTTTAAATTTTTT | -3201 |
| TCCAGCAATCCTCCTGCCTCAGCCTCCCAGAGTATTGGGATTACAGGTGTGAGCCATTGTGCTTGATCAAGATGCTGTTATGGGCTGAGTTGTGTTCCTC | -3101 |
| AAAAATTCTCTTGAAGTCCTAATCTCAAGTACTTCAGGACGTGACCTTATTTGAAGGACCCCCCTTATAGGGTCTTTACAGAGGTAATTAAGTTAAAATG | -3001 |
| AGGCCATTAGGATGGGGCCTAATGCAATATGACTGGTATCCTTGAAAAAAGGGGAAACTTGGAGACTGACT | -2901 |
| GAAAATGGCCAAGGAGGGGGGGGGCCTGGAATAGAGCCTTCCTT | -2801 |
| AGCCTCCAGCATCTTGAGAGATTTCTGTTGTTTAAGTCATGCAATATGTAGTACTTTGTTACAGCAGCCCTAGCAAACTGATACACTCACCAAATCGATT | -2701 |
| TTGTGACTCACTATTGGGTTGTAACCAGCAGTACATAGACATAAAGTTATTTTTCCTTACGCTTTATCTTGTGCAATCGTGTGTGT | -2601 |
| TGTGTGTGTGTGTGTGTGTGTGTGACGGAGTCTTGTTCTGTCACCAGGCTGGAGTGCAGTGGCTTGATCTCGGCTCACTATAATCACAGCCTTCCAGATT Bss HII | -2501 |
| CAAGTGATTTCCCTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGGCGCGCACCACCACGCCCGACTAATTTTTTGTATTTTAGTAGAGACGGGGTTTC | -2401 |
| ACCATGTTGGCCAGGATGGTCTCAATCTCCTGACCTTGTGATCTGCCTGC | -2301 |
| TCTTTACCACCACTCAATGGGATGTCAAGGTCCAGGGGGGGG | -2201 |
| ACCCTGCAATTGTACAATCTGAAGATGAGTATCTCCTTAAATTTCATACTCTAGGCACTTTACCCTAGCCTAGACTCTGTTGAAGTAGGTATAACTATTA AP-2 | -2101 |
| TTCTCATTTGAGGGATTGACACCTGATTGTGAACCTCCTAAATGGAGTCATACCCAAGCCAGATTTGCCTCTAAATTCTGTTTTTTCCCCTTACATCACA C/EBP-δ | -2001 |
| GTGTTCCCATTGGTATAGTCAGTTACAGAGGGAGTAATATATACTATTTTTCTACCAGTACTTGCTCCTCGCCTTCCTACCCCCTAAAAAGGAGCCAAAGT | -1901 |
| CAGAGATCACATTTACTCTTTTCCCTCCTCCTCCCAAGTCTTTGGGGACTTGTAGCTCTGACACCCTTAGATGGTGAAACCTGGCTTCACCTACTGTCT | -1801 |
| GTGGATGTCTGCAGGCAGAGTGGGCACTCAGGAGCACATACAAAGCACGTGTGCCGTGAACACGTATGTGCACACACCTTGATCCTAGCATGGCTTGTTG CCAAT Enhancer | -1701 |
| GACAAGCCAATGGACAGAGTCCCTGCCTGCCACCCCCCCC | -1601 |
| CACCCTCAATATCCTTTTGGCAGTAACATGACCCCCCAAATCTGGGGACTTCTATGTAGGATGGAGACCCTTCTCCTCATACCTGGTTTATTATGA | -1501 |
| $\begin{array}{c} \text{ACCATAAAAATAGTGCCTGACAGTTACTGTGTGTGTGCAGGCATTGTTCTAAGCCTTCAGATGTTTTACTGCATTTTATTCTCACATTATGGGTTAAGACTTA \\ \text{ AP-1 } & \text{ AP-3 } \end{array}$ | -1401 |
| TTTGCTCCATTTTACAGATGACGAGAATGAATCACAGAGTAAATTGCTCAGGGTTGTGTGGTTAGCAGCATTAGCAGGATTTGAACCCAAGCAGCATGTA | -1301 |
| TCCACAGTCCAGTCTTTTAACTGCTATATTTTGCTGTGTTCAAACCCTCTGCTGCCTGGCTGG | -1201 |
| $\begin{array}{c} Agcaaggaggaggaggaggaggaggaggaggaggaggagga$ | -1101 |
| GGAGGGAGGAGGAGGAGGAGTAGAAACTTCCTCCCACCGCCCCTCAGACACCACCTCTTCCACACACCGGGGCTCTCAGGTGTCCGGGAGTAAAGGCCTCTCTG <u>Spel</u> <u>AP-1</u> | -1001 |
| GATCCCTTGGTCTCCTCCAGCTCCTCCCCCAGCAAAAACTGCAGAACCCTCCACTAGTTATGTTGATGACTCAGAAGTTGAGCAAGACTGTGTGTG | - 901 |
| TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG | - 801 |
| $\label{eq:construct} TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG$ | - 701 |
| $\begin{array}{c} ATAGGAACACAGCAGATTCGTATTCAAACTTGCCCCTTGTGAATCTGCAGGCAG$ | - 601 |
| $\hline CCTGTGAGGAGAGGAGAGAGCAAAGACCAGCTTCAGTCCAAGGGACTCCTAGAGTCTTCCAGAATTCTGAGCTGAGGGTTCCCCTCCCCCACTCCCGTCAGBInI$ | - 501 |
| | - 401 |
| TGGTCACGAGACCGACCTCTAAGGCGTTCCCCIGCCGGAAGGGAGGGGGGGGGG | |
| $\frac{Bsu36 \ I}{CTGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG$ | - 301 |
| $\begin{tabular}{lllllllllllllllllllllllllllllllllll$ | - 301 - 201 |
| $\begin{tabular}{lllll} \hline TGGTCACGAGACCGACCTCTAAGGCGTTCCCTGCCGGAAGGCAGGGGGGACCAGGAGTGGAGTTGGCTGGC$ | - 301 - 201 - 101 |
| $\begin{tabular}{lllllllllllllllllllllllllllllllllll$ | - 301 - 201 - 101 - 1 |
| $\begin{tabular}{lllllllllllllllllllllllllllllllllll$ | - 301 - 201 - 101 - 1 + 100 |

FIG. 3. The 5'-flanking sequence of the human Stat6 gene. The sequence of the putative proximal promoter (uppercase letters) including 134 bp of exon 1 (lowercase letters) of the human Stat6 gene is shown, with +1 corresponding to a prominent upstream transcriptional initiation site. Additional potential initiation sites are indicated by asterisks (*). Potential transcriptional elements and restriction endonuclease cleavage sites are indicated.

AUG translation initiation codon encoding the first amino acid was located within exon 3. The DNA binding domain and the putative SH3-like domain were transcribed from exons 10–14 and 14–15, respectively. Most of the SH2 domain, except the first four and last three codons, was contained within exon 16 and 17. The phosphorylated tyrosine residue of Stat6 (Y641) critical for SH2 domain interaction was encoded within exon 17. Two regions of maximal amino acid homology conserved among Stat family members were distributed within exons 10 and 13.

Characterization of the Stat6 Gene Start Site and 5' Flanking Sequence

To determine the origin of Stat6 transcription, RNase protection and primer extension experiments were performed. Hybridization of an antisense riboprobe generated from a 497-bp Bsu36I-NheI genomic fragment with human Ramos poly(A)⁺ RNA resulted in a protected band of 134 bases (Fig. 2) that corresponded to our longest cDNA clone. Several protected bands downstream of this species and thus located within the 5' noncoding regions of our cDNA clones (107 and 97 bases) were also detected (Fig. 2). A minor species of 190 bases was also observed. Taken together, these results suggest multiple origins of transcription perhaps due to variant messages. Primer extension experiments proved to be difficult, most likely due to the high GC content of the template. Based on RNase protection assay and our longest cDNA, we have tentatively assigned a Stat6 mRNA start site (marked +1 in Fig. 3) 318 bp upstream of the AUG translation initiation codon.

To characterize further the putative promoter, restriction mapping of the human Stat6 λ 3 clone using single or double restriction endonuclease digestions combined with Southern blot hybridization identified an 8.2-kb *KpnI–Sal*I fragment that contained a 5'flanking sequence and exons 1–3. The 8.2-kb *KpnI–Sal*I fragment was subcloned and sequenced. The nucleotide sequence of approximately 3.6 kb of the human Stat6 5' flanking sequence and 134 bp of 5' untranslated sequence contained within the Stat6 cDNA clones is shown in Fig. 3.

To identify potential transcription factor consensus binding elements, transcription factor database 7.3 (IntelliGenetics, Inc., Mountain View, CA) was used to perform computer analysis of the Stat6 5' flanking/putative promoter sequence. As shown in Fig. 3, the promoter contained one GC box, the proposed site for binding Sp1 transcription factor. As observed with many myeloid promoters, the Stat6 promoter contained neither a TATA box nor a CCAAT sequence upstream (within 100 bp) of the transcription start site. However, the Stat6 5' flanking sequence contained a TATA box and CCAAT sequence at -2522 and -1694, respectively (Fig. 3). Several transcriptional regulatory consensus sequences were also found within the putative promoter. In addition to



FIG. 4. Deletion analysis of basal promoter activity of the human Stat6 gene. Varying lengths of Stat6 upstream flanking sequence (approximately -5500 to +134) were fused to a pGL3-Basic luciferase reporter vector, and luciferase activity was determined. Black and open boxes represent regions of the Stat6 gene 5'-flanking sequence and the luciferase reporter gene, respectively. Restriction endonuclease cleavage sites depicted are identical to those indicated in Fig. 3. The relative luciferase activity is shown with the values determined after normalization as described under Materials and Methods.

one binding site each for transcription factors III-A, SIF, and C/EBP- δ , there are two sites for NF-IL6 and NF κ B, seven sites for GMCSF, nine sites for γ -IRE, two sites for activator protein (AP)-1, eight sites for AP-2, and one site for AP-3.

Functional Analysis of Human Stat6 Promoter

To confirm that the isolated *Kpn*I–*Sal*I DNA fragment indeed contained a functional promoter, we prepared reporter gene constructs by inserting the 5'-flanking region of human Stat6 as a *Kpn*I–*Nhe*I fragment (approximately -5500 bp to +134 bp, designated pBP78) upstream of the firefly luciferase gene in the pGL3-Basic vector. To identify the minimal promoter region required for the maximal constitutive reporter activity, sequential 5'-deletion constructs of pBP78 were engineered (Fig. 4). After transient transfection of NIH 3T3 cells, we observed a significant level of transcriptional activity with several constructs (Fig. 4). Maximal basal luciferase activity was observed with pBP86 (-947 to +134). Subsequent deletion of pBP86 resulted in reduced luciferase activity. The parental vector produced only trace amounts of transcriptional activity. Similar



FIG. 5. FISH localization of Stat6 gene to human chromosomes. Digital image of a metaphase chromosome spread derived from methotrexate-synchronized normal human peripheral leukocytes after hybridization with a digoxigenin-labeled/genomic fragment probe and DAPI counterstaining. Both chromosomes with symmetrical rhodamine signals on sister chromatids were identified as chromosome 12, and the signal was localized at bands 12q13.3–q14.1.

results were obtained when these constructs were transfected into HepG2 cells (data not shown). These results established that the 5'-flanking region of the Stat6 gene contained a promoter that is constitutively active in fibroblast and epithelial cell types.

Chromosomal Localization of the Human Stat6 Gene

The chromosomal location of the human Stat6 gene was determined to gain further clues regarding function as well as to investigate potential linkage with disease. In two FISH experiments with lymphocytes from different individuals, the majority of the chromosomes had specific fluorescent signals at identical sites on both chromatids of chromosome 12. From a total of 100 metaphases examined, 90 had fluorescent signals on chromosome 12 and 85 had both homologues of chromosome 12 labeled. Double fluorescent labels were not observed at any other site and the five single signals also detected were randomly distributed over different chromosomes. A single specific site of hybridization for the gene unequivocally demonstrates a single locus for this gene. The location of the fluorescent signal was determined directly in 50 metaphases with DAPI-enhanced G-like banding at region 12q13.3-q14.1, where we assigned the locus of the human Stat6 gene (Fig. 5).

DISCUSSION

Here we describe the human Stat6 gene exonic structure, putative promoter, and chromosomal location. The human Stat6 gene extended over 19 kb and consisted of 23 exons and 22 introns. The 5'-flanking region that contained the putative promoter region was also characterized, and several potential regulatory elements were identified. By fluorescence *in situ* hybridization, the human Stat6 gene was mapped to chromosome 12q13.3–q14.1. This location was further confirmed by the partial coding sequence of the NAB2 gene, which was recently mapped to an identical locus (Svaren *et al.*, 1996, 1997) and found within the distal 3'-flanking region of our human Stat6 λ 4 genomic clone (B. K. R. Patel and W. J. LaRochelle, unpublished observation).

Among the human Stat genes, only the exonic structures of Stat1 and Stat2 have been determined. Both the Stat1 and the Stat2 genes comprised 24 exons (Yan et al., 1995), while Stat6 contained 23 exons. The Stat1 and 2 exonic division divided both these proteins at almost identical sites (Yan et al., 1995). The relationship between the genomic structure and the encoded functional domains of Stat1 and Stat2 was conserved only somewhat compared to Stat6. In particular, the proposed SH3 domain of Stat1 and 2, like that of Stat6, was encoded by three exons. The tyrosine residue phosphorylated in Stat1 and 2, analogous to that of Stat6 (Y641), was also encoded within a separate exon. However, the Stat1 and 2 SH2 domain was encoded by three (Yan et al., 1995), rather than two, exons as determined for Stat6. Additionally, while a few variations were noted to the GT/AG rule for Stat1 and 2



FIG. 6. Alternative splicing of Stat6 variant mRNAs. The Stat6 genomic DNA sequences around exon (uppercase letters)–intron (lowercase letters) junctions potentially affected by alternative splicing are presented. The resulting Stat6, Stat6b, and Stat6c variant mRNAs due to the proposed splicing of exon 4 (**A**), exon 2 (**B**) and exon 16 (**C**) are shown.

exon-intron boundaries, no exceptions were observed for Stat6.

We have recently isolated three Stat6 cDNAs, including wildtype Stat6 as well as the attenuated Stat6b and dominant negative Stat6c variants (Patel et al., 1998). The transcripts of all three variants were expressed to varying extents in vivo. The three human Stat6 variants appear to arise from differential splicing that was best explained by selective exon deletion and utilization of alternative donor/acceptor sites (Fig. 6). The wildtype Stat6 cDNA resulted from the deletion of exon 2 apparently by ignoring splice donor and acceptor sites at the beginning and end of exon 2. Similarly, the Stat6b variant resulted from a splicing event that deleted exon 4 in addition to exon 2. The differential splicing that gave rise to Stat6c retained exon 2 by recognition of these donor and acceptor sites. Stat6c apparently also deleted an 84-bp region of exon 16 utilizing an alternative donor/acceptor splice site within exon 16. Interestingly, only Stat6 has been found to utilize alternative donor/acceptor sites within SH2 domain-encoded sequences to produce a dominant

negative variant, although all other Stat family members possess some form of differential splicing.

The chromosomal location of the human Stat6 gene was determined so that linkage with a possible disease locus could be established. In this regard, the human Stat6 gene was mapped to chromosome 12q13.3-q14.1, consistent with that predicted from the known mouse chromosomal location (Copeland et al., 1995; Svaren et al., 1997). This locus contained the ErbB-3 and Gli proto-oncogenes (Kraus et al., 1989), the RAP1B tumor suppressor (Kitamaya et al., 1989), and homeobox as well as keratin gene loci (Cannizzaro et al., 1987; Popescu et al., 1987). Multiple aberrancies about this locus have been found (Van de Ven et al., 1995). Benign solid tumors that often display chromosomal breakpoints in the 12q13-q15 region include uterine leiomyoma, adenoma, hamartoma, chondroma, and endometrial polyps (Solomon et al., 1991). Additionally, 12q13-q15 has also been found to be a human papilloma virus-18 integration site in cervical carcinoma (Popescu et al., 1987; Sastre-Garau et al., 1990). 12q13 translocations have been reported in mesenchymal tumors such as myxoid liposarcomas, round cell liposarcomas, clear-cell sarcomas, and rhabdomyosarcomas, some of which were attributed to TLS-CHOP t(12; 16)(q13;p11) (Dal Cin *et al.*, 1997; Solomon *et al.*, 1991). Amplification of MDM2, CDK4, and SAS (12q13–q14) was often observed in human sarcomas and glioblastomas (Meltzer *et al.*, 1991; Oliner *et al.*, 1992; Smith *et al.*, 1992). The contribution of Stat6 to these malignancies remains to be investigated.

ACKNOWLEDGMENTS

We thank Drs. Jackie Pierce, Matt Lorenzi, Prateek Jindal, and Stefanie Nelson for helpful discussions. We also acknowledge Paul Kriebel, Rayah Blumenthal, and Veena Kapoor for valuable technical assistance.

REFERENCES

- Aaronson, S. A. (1991). Growth factors and cancer. Science 254: 1146-1153.
- Beeler, J. F., Patel, B. K. R., Chedid, M., and LaRochelle, W. J. (1997). Cloning and characterization of the mouse homolog of the human A6 gene. *Gene* 193: 31–37.
- Cannizzaro, L. A., Croce, C. M., Griffin, C. A., Simeone, A., Bonicinelli, E., and Huebner, K. (1987). Human homeo box-containing genes located at chromosome regions 2q31–2q37 and 12q12– 12q13. Am. J. Hum. Genet. 41: 1–15.
- Copeland, N. G., Gilbert, D. J., Schindler, C., Zhong, Z., Wen, Z., Darnell, J. J. E., Mui, A. L. F., Miyajima, A., Quelle, F. W., Ihle, J. N., and Jenkins, N. A. (1995). Distribution of the mammalian Stat gene family in mouse chromosomes. *Genomics* **29**: 225–228.
- Dal Cin, P., Sciot, R., Panagopoulos, I., Aman, P., Samson, I., Mandahl, N., Mitelman, F., Van den Berghe, H., and Fletcher, C. D. (1997). Additional evidence of a variant translocation t(12;22) with EWS/CHOP fusion in myxoid liposarcoma: Clinicopathological features. J. Pathol. 182: 437–441.
- Darnell, J. E., Jr., Kerr, I. M., and Stark, G. R. (1994). Jak–STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* 264: 1415–1421.
- Fleming, T. P., Matsui, T., Heidaran, M. A., Molloy, C. J., Artrip, J., and Aaronson, S. A. (1992). Amplification and/or overexpression of platelet-derived growth factor receptors and epidermal growth factor receptor in human glial tumors. *Cancer Res.* 52: 4550–4553.
- Heldin, C.-H., and Westermark, B. (1991). Platelet-derived growth factor and autocrine mechanisms of oncogenic processes. *Crit. Rev. Oncog.* **2:** 109–124.
- Hou, J., Schindler, U., Henzel, W. J., Ho, T. C., Brasseur, M., and McKnight, S. L. (1994). An interleukin-4-induced transcription factor: IL-4 Stat. *Science* 165: 1701–1706.
- Ihle, J. N., and Kerr, I.M. (1995). Jaks and Stats in signaling by the cytokine receptor superfamily. *Trends Genet.* **11**: 3951–3963.
- Kaplan, M. H., Schindler, U., Smiley, S. T., and Grusby, M. J. (1996). Stat6 is required for mediating responses to IL-4 and for the development if Th2 cells. *Immunity* 4: 1–20.
- Kitamaya, H., Sugimoto, Y., Matsuzaki, T., Ikawa, Y., and Nada, M. (1989). A ras-related gene with transformation suppressor activity. *Cell* **56**: 77–84.
- Kraus, M. H., Issing, W., Miki, T., Popescu, N. C., and Aaronson, S. A. (1989). Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor family: Evidences for overexpression in a subset of human mammary tumors. *Proc. Natl. Acad. Sci. USA* 86: 9193–9197.
- Malabarba, M. G., Rui, H., Deutsch, H. H., Chung, J., Kalthoff, F. S., Farrar, W. L., and Kirken, R. A. (1996). Interleukin-13 is a potent

activator of JAK3 and STAT6 in cells expressing interleukin-2 receptor-gamma and interleukin-4 receptor alpha. *Biochem. J.* **319:** 865–872.

- Meltzer, P. S., Jankowski, S. A., Dal Cin, P., Sandberg, A. A., Paz, B., and Coccia, M. A. (1991). Identification and cloning of a novel amplified sequence in human malignant fibrous histiocytoma derived from a region of chromosome 12 frequently rearranged in soft tissue tumors. *Cell Growth Differ.* 2: 495–501.
- Nister, M., Claesson-Welsh, L., Eriksson, A., Heldin, C.-H., and Westermark, B. (1991). Differential expression of platelet-derived growth factor receptor in human malignant glioma cell lines. *J. Biol. Chem.* 266: 16755–16763.
- Oliner, J. D., Kinzler, K. W., Meltzer, P. S., George, D. L., and Vogelstein, B. (1992). Amplification of a gene encoding a p53associated protein in human sarcomas. *Nature* **253**: 80–83.
- Patel, B. K. R., Pierce, J. H., and LaRochelle, W. J. (1998). Regulation of interleukin-4 mediated signaling by naturally occurring dominant negative and attenuated forms of human Stat6. *Proc. Natl. Acad. Sci. USA* 95: 172–177.
- Patel, B. K. R., Wang, L. M., Lee, C. C., Taylor, W. G., Pierce, J. H., and LaRochelle, W. J. (1996). Stat6 and Jak1 are common elements in platelet-derived growth factor and interleukin-4 signal transduction pathways in NIH 3T3 fibroblasts. *J. Biol. Chem.* 21: 22175–22182.
- Pernis, A., Witthuhn, B., Keegan, A. D., Nelms, K., Garfein, E., Ihle, J. N., Paul, W. E., Pierce, J. H., and Rothman, P. (1995). Interleukin 4 signals through two related pathways. *Proc. Natl. Acad. Sci.* USA 92: 7971–7975.
- Popescu, N. C., Amsbaugh, S. C., and DiPaolo, J. A. (1987). Human papillomavirus type 18 DNA is integrated at a single chromosome site in cervical carcinoma cell line SW576. *J. Virol.* 61: 1682–1685.
- Popescu, N. C., Zimonjic, D., Hatch, C., and Bonner, W. (1994). Chromosomal mapping of the human histone gene H2AZ to 4q24 by fluorescence *in situ* hybridization. *Genomics* **20**: 333–335.
- Quelle, F. W., Shimoda, K., Thierfelder, W., Fischer, C., Kim, A., Ruben, S. M., Cleveland, J. L., Pierce, J. H., Keegan, A. D., Nelms, K., Paul, W. E., and Ihle, J. N. (1995). Cloning of murine Stat6 and human Stat6, Stat proteins that are tyrosine phosphorylated in responses to IL-4 and IL-3 but are not required for mitogenesis. *Mol. Cell. Biol.* 15: 3336–3343.
- Sastre-Garau, X., Schneider-Maunoury, S., Couturier, J., and Orth, G. (1990). Human papillomavirus type 16 DNA is integrated in chromosome region 12q14-q15 in a cell line derived from a vulvar intraepithelial neoplasia. *Cancer Genet. Cyto*genet. 44: 253-261.
- Schindler, C., and Darnell, J. E., Jr. (1995). Transcriptional responses to polypeptide ligands: The JAK–STAT pathway. *Annu. Rev.Biochem.* 64: 621–651.
- Schindler, C., Kashleva, H., Pernis, A., Pine, R., and Rothman, P. (1994). STF-IL-4: A novel IL-4-induced signal transducing factor. *EMBO J.* 13: 1350–1356.
- Schindler, U., Wu, P., Rothe, M., Brasseur, M., and McKnight, S. L. (1995). Components of a Stat recognition code: Evidence for two layers of molecular selectivity. *Immunity* 2: 689–697.
- Seder, R. A., Paul, W. E., Davis, M. M., and Fazekas de St. Groth, B. (1992). The presence of interleukin 4 during in vitro priming determines the lymphokine-producing potential of CD4⁺ T cells from T cell receptor transgenic mice. *J. Exp. Med.* **176**: 1091–1098.
- Shimoda, K., Deursen, J., Sangster, M. Y., Sarawar, S. R., Carson, R. T., Tripp, R. A., Chu, C., Quelle, F. W., Nosaka, T., Vignali, D. A. A., Doherty, P. C., Grosveld, G., Paul, W. E., and Ihle, J. N. (1996). Lack of IL-4-induced Th2 responses and IgE class switching in mice with disrupted Stat6 gene. *Nature* **380**: 630– 633.
- Shuai, K., Horvath, C. M., Huang, L. H. T., Qureshi, S., Cowburn, D., and Darnell, J. E., Jr. (1994). Interferon activation of the tran-

scription factor Stat91 involves dimerization through SH2-phosphotyrosyl peptide interactions. *Cell* **76**: 821–828.

- Smith, S. H., Weiss, S. W., Jankowski, S. A., Coccia, M. A., and Meltzer, P. S. (1992). SAS amplification in soft tissue sarcoma. *Cancer Res.* 53: 3746–3749.
- Smits, A., Funa, K., Vassbotn, F. S., Beausang-Linder, M., af Ekenstam, F., Heldin, C.-H., Westermark, B., and Nister, M. (1992). Expression of platelet-derived growth factor and its receptors in proliferative disorders of fibroblastic origin. *Am. J. Pathol.* 140: 639–648.
- Solomon, E., Borrow, J., and Goddard, A. D. (1991). Chromosome aberrations and cancer. *Science* **254**: 1153–1160.
- Svaren, J., Apel, E. D., Simburger, K. S., Jenkins, N. A., Gilbert, D. J., Copeland, N. A., and Milbrandt, J. (1997). The Nab2 and Stat6 genes share a common transcription termination region. *Genomics* **41**: 33–39.
- Svaren, J., Sevetson, B. R., Apel, E. D., Zimonjic, D. B., Popescu, N. C., and Milbrandt, J. (1996). NAB2, a corepressor of NGFI-A (Egr-1) and Krox20, is induced by proliferative and differentiative stimuli. *Mol. Cell. Biol.* 16: 3545–3553.

- Takeda, K., Tanaka, T., Shi, W., Matsumoto, M., Minami, M., Kashiwamura, S., Nakanishi, K., Yoshida, N., Kishimoto, T., and Akira, S. (1996). Essential role of Stat6 in IL-4 signalling. *Nature* 380: 627–630.
- Van de Ven, W. J. M., Schoenmakers, E. F. P. M., Wanschura, S., Kazmierczak, B., Kools, P. F. J., Geurts, J. M. W., Bartnitzke, S., Van den Berghe, H., and Bullerdiek, J. (1995). Molecular characterization of MAR, a multiple aberration region on human chromosome segment 12q13–q15 implicated in various solid tumors. *Genes Chromosomes Cancer* 12: 296–303.
- Yan, R., Qureshi, S., Zhong, Z., Wen, Z., and Darnell, J. E., Jr. (1995). The genomic structure of the STAT genes: Multiple exons in coincident sites in Stat1 and Stat2. *Nucleic Acids Res.* 23: 459-463.
- Zimonjic, D. B., Rezanka, L., DiPaolo, J. A., and Popescu, N. C. (1995). Refined localization of the erbB-3 proto-oncogene by direct visualization of FISH signals on LUT-inverted and contrastenhanced digital images of DAPI-banded chromosomes. *Cancer Genet. Cytogenet.* 80: 100–102.